

# Machine Translation - Mystery, Misery or Miracle

**Fred Klein**



As one of the first users of commercial MT in the United States, and as a senior professional translator, I see MT as one of many "tools". As an independent expert without connections to the industry I can be objective. Since 1980 I have used one system for years and have worked on and tested others. Few translators have years of experience in both the conventional and the MT fields.

MT -- as opposed to Computer Assisted Translation uses the computer (the machine) to perform a total translation online. CAT uses the computer as a tool to assist the human translator.

We do not know how we think. Only part of the human translation is understood and "programmable". Concrete, defined grammars elude us and language is constantly changing, living and expanding. These are not the best bases for programmers and computational linguists. The brain is and will be the unsurpassed "computer" in humans.

## History

Fifty years ago, MT began in England and the Cold War promoted it.

In the United States, the Defense Department needed translations of millions of words from Russian *now*. A Hungarian, Dr. Toma, was the right man at the right time: he bragged of a 300,000 word output per hour on a mainframe. There was no quality control - it was all confidential. Another company, Logos, translated Vietnamese. The keys were online bilingual glossaries, over one million terms in the case of Toma. Taxpayers paid for the enormous investment.

Later, the Mormon Church tried MT and abandoned it. The European Community uses a minor part of translations by MT (Toma's Systran).

Some private corporations have tried to profit from MT in the United States, in Europe (including Russia), and in Japan above all. I was told by the leading Japanese MT expert that there are so few human translators of Japanese that they must have MT.

## The Crux of the Matter (Machine Translation, that is)

A machine cannot think (forget artificial intelligence for now).

MT requires an online bilingual glossary, a transfer mechanism, and human post-editing as a minimum. If you happen to need 30-year-old terms from Russian to English (updated in part), you are lucky because a large glossary already exists

(Systran). In other languages, vendors offer general glossaries of from 5,000 to 100,000 words, but it is unlikely that your particular technical vocabulary exists online in the language *you* need.

So you run a count of words not found. A document with a corpus of 30,000 (the total amount of words) may include 14,000 words that are not found. Every word not found, including proper names, misspelled words, unknown acronyms, even verbs and pronouns, has to be researched (in dictionaries), input, coded and tested. The amount of time and money needed is prohibitive. Remember, the machine has rules, but does not know any words other than those in the original online glossary!

The transfer mechanism -- grammar, word position etc. -- is not as critical and sometimes quite good. Even post-editing is tolerable.

### **Probability, Statistics, Predictability - the Essential Elements**

Assuming we work in private industry, the investment in one document is unreasonable. Assuming there is a guarantee that in the next three years there will be 300 documents with a similar vocabulary, then the ratio of investment to output would be favorable.

But, generally speaking, the private sector cannot predict such similar documents because the market and technology keep changing.

### **The Public Sector: Limits, No Competition or Profit Margins**

Take the field of weather forecasting. The TAUM (Automatic Translation System of the University of Montreal) is a classic example. Meteorology is stable, limited and predictable. The French forecast follows the English one throughout Canada in 20 minutes, and almost without human intervention. A private company has contracted with the government. Great!!

The problem is hidden. The government asked the TAUM group to design a similar system for aviation hydraulics. After 3 years, the MT people gave up. Human translation was faster and cheaper!! Even this limited field was too complicated.

Public sectors offer some great opportunities: The PanAmerican Health Organization has a first class MT system, called SpanAm (Spanish American). One language, Spanish, is spoken all over Latin America (except in Brasil). Health terms are relatively stable and documents repeat year after year. An outstanding example. But the technology is not available to outsiders.

There have been disasters, like the "Eurotra" project for Europe!!

### **The Private Sector**

Companies like Unisys, Xerox, Caterpillar, Siemens and John Deere have tried MT, but these are only a few compared to the thousands of global corporations. My recommendation is to stick with all kinds of CAT systems. There is little chance

that private translation agencies could use MT. Think of consistent terms, fuzzy memory and the likes where the computer is your assistant, not your master.

In testing MT systems, beware of "prototypes" and edited texts. Sit sit down at the keyboard, try it yourself, and beware of salespeople!